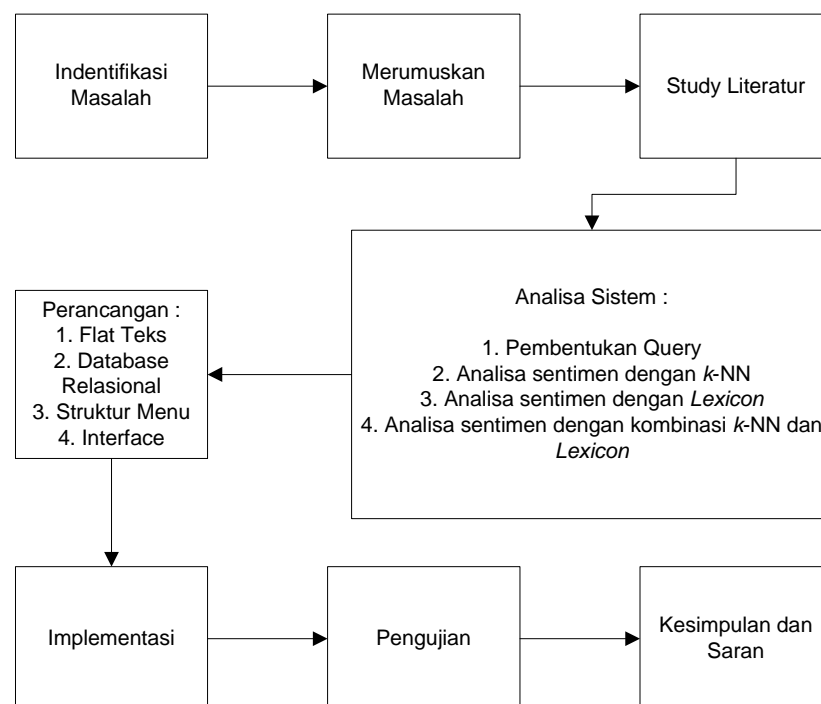


## BAB III

### METODOLOGI PENELITIAN

Tahapan penelitian yang dilakukan dalam penelitian adalah sebagai berikut.



**Gambar 3.1 Tahapan Penelitian**

Berdasarkan Gambar 3.1 terdapat 8 tahapan dalam penelitian ini. Akan dijelaskan secara rinci tahapan pada penulisan dibawah ini.

#### 3.1 Identifikasi Masalah

Pengguna Twitter dapat menuliskan pesan apapun pada Twitter. Sehingga membuat beragamnya topik yang dapat ditemukan pada situs *microbloging* ini. Seringkali terdapat kata negasi dalam kalimat sentimen. Disamping itu pula cakupan penulisan pun terbatas yaitu 140 karakter. Terbatasnya penulisan

membuat pengguna Twitter menuliskan pesan berupa singkatan. Pengguna Twitter juga dapat mengungkapkan ekspresi mereka dengan penggunaan huruf berlebihan pada pesan Twitter. Hal ini menjadi permasalahan sendiri dalam menemukan orientasi sentimen analisis pada Twitter. Untuk itu diperlukan Algoritma yang mampu menemukan pesan Twitter yang mengandung sentimen sekaligus dapat menyeleksi kata slang dan kemudian merubahnya dengan data yang dapat dijadikan fitur.

### **3.2 Merumuskan Masalah**

Dari tahapan identifikasi masalah dapat dirumuskan bagaimana menemukan pesan Twitter yang mengandung sentimen sekaligus dapat menyeleksi kata slang dan kemudian merubahnya dengan data yang dapat dijadikan fitur.

### **3.3 Study Literatur**

Pada tahap ini akan dilakukan pengumpulan informasi dan menemukan referensi yang berhubungan dengan permasalahan pada penelitian ini melalui jurnal ilmiah dan buku terkait.

### **3.4 Analisa Aplikasi**

Pada tahapan ini dijelaskan secara rinci tentang proses sistem yang dibangun, sehingga mempermudah pemahaman terhadap sistem. Dalam penelitian ini akan ditentukan analisis sentimen dengan menggunakan tahapan kombinasi algoritma *k*-NN dan pendekatan *lexicon*. Disamping itu sebagai perbandingan dalam penelitian ini juga dibangun penentuan analisis sentimen dengan tahapan *k*-NN dan pendekatan *lexicon* secara terpisah. Pada tahap ini terbagi menjadi beberapa proses :

- a. Analisa penentuan pesan Twitter yang akan dijadikan *query* untuk ditentukan sentimen analisisnya. Pada tahap ini terdapat beberapa langkah diantaranya :
  - a. Seleksi Tweet berita : Tweet berita mengandung URL dari alamat berita. Tweet berita merupakan fakta dimana tidak dapat ditentukan sentimen.

- b. Seleksi Duplikat Tweet : Tweet duplikat terjadi dikarenakan adanya *spamer*.
- c. Seleksi Tweet RT : Tweet RT merupakan tweet terusan dari tweet asal.
- b. Analisa Sistem orientasi sentimen pada *query* menggunakan Algoritma *k*-NN. Tahapan ini terdiri dari beberapa langkah, yakni
  - a. Pembangunan index (pembuatan *inverted index*) dan pembobotan kata pada data latih Algoritma K-NN menggunakan teknik TF-IDF. Pada tahap ini akan dilakukan :
    1. Mengumpulkan dokumen yang didapat dari kumpulan tweet berisi sentimen sebagai data latih, kemudian memberikan label sentimen positif atau negatif setiap dokumen secara manual. Proses selanjutnya akan dilakukan melalui sistem.
    2. *Case folding* : tahapan ini merubah semua huruf kapital dengan huruf kecil.
    3. Menghapus karakter selain UTF-8, link URL dan dokumen yang berulang
    4. Pemisahan rangkaian kata (*tokenization*).
    5. Melakukan *linguistic preprocessing*: *stopword* dan *stemming*
    6. Normalisasi : tahapan identifikasi kata slang dan penulisan kata berlebihan kemudian di ganti dengan kata kamus KBBI
    7. *Indexing*.
    8. Pembobotan kata.
  - b. *Preprocessing query*. Tahapan yang dilakukan adalah :
 

*Case folding*, menghapus karakter selain UTF-8, link URL dan dokumen yang berulang, *tokenization*, melakukan *linguistic preprocessing*, normalisasi dan Pembobotan kata.
  - c. Klasifikasi *query* dengan algoritma *k*-NN. Tahapan yang dilakukan
    1. Menghitung kedekatan kemiripan query dengan persamaan *Cosine Similarity* dengan persamaan 2.5 pada data latih.
    2. Menentukan nilai *k* pada algoritma *k*-NN.
    3. Menyimpulkan orientasi sentimen dengan persamaan 2.6.

- c. Analisa Sistem orientasi sentimen pada *query* menggunakan pendekatan *Lexicon*. Tahapan ini terbagi menjadi 2 langkah, yakni
  - a. Membangun kamus sentimen yang terbagi menjadi sentimen positif dan negatif.
  - b. Membangun kamus negasi.
  - c. Menentukan orientasi sentimen pada *query*. Tahapan yang dilakukan adalah:
    1. Identifikasi setiap kata sentimen pada *query*.
    2. Identifikasi kalimat negasi.
    3. Identifikasi klausa Tapi-
    4. Hitung score dengan persamaan 2.1
- d. Analisa Sistem orientasi sentimen pada *query* menggunakan kombinasi Algoritma *k*-NN dan *Lexicon*. Tahapan ini terdiri dari beberapa langkah, yakni
  - Kombinasi A
    - a. Membangun kamus sentimen yang terbagi menjadi sentimen positif dan negatif.
    - b. Pembangunan index (pembuatan *inverted index*) dan pembobotan kata pada data latih Algoritma *k*-NN menggunakan teknik TF-IDF.
  - c. *Preprocessing query*. Tahapan yang dilakukan adalah :
    1. *Case folding*
    2. Menghapus karakter selain UTF, link URL dan dokumen yang berulang
    3. Pemisahan rangkaian kata (tokenization).
    4. Melakukan *linguistic preprocessing: stopword* dan *stemming*
    5. Normalisasi
    6. Seleksi kalimat netral menggunakan pendekatan *Lexicon*
    7. Pembobotan kata.
  - d. Klasifikasi *query* dengan algoritma *k*-NN. Tahapan yang dilakukan
    1. Menentukan nilai *k* pada algoritma *k*-NN.
    2. Menghitung kedekatan kemiripan *query* dengan persamaan *Cosine Similarity* dengan persamaan 2.5 pada data latih.
    3. Menyimpulkan orientasi sentimen sementara dengan persamaan 2.6

- e. Evaluasi orientasi sementara dengan pendekatan *Lexicon*
  - 1. Identifikasi kata kunci menggunakan kamus sentimen
  - 2. Identifikasi kalimat negasi.
  - 3. Identifikasi klausa Tapi-
  - 4. Tentukan orientasi akhir sentimen.
- Kombinasi B
  - a. Membangun kamus sentimen yang terbagi menjadi sentimen positif dan negatif.
  - b. Membangun kamus negasi.
  - c. Menentukan orientasi sentimen dengan metode *lexicon*.
    - 1. Identifikasi kata sentiment. Jika tidak ditemukan maka orientasi bernilai netral.
    - 2. Jika Ada, namun diidentifikasi kata sentiment bernilai seimbang antara kata positif dan negatif, proses berlanjut dengan menentukan orientasi dengan metode k-NN. Jika tidak memiliki sentiment seimbang hasil.

### 3.5 Perancangan Aplikasi

Perancangan berarti metode yang khusus digunakan untuk merancang sistem yang telah dianalisa dengan tujuan untuk memberikan kemudahan dan menyederhanakan suatu proses atau jalannya aliran data, perancangan terhadap model, dan merancang rancang bangun sistem. Adapun rancangan utama sistem yaitu :

- 1. Perancangan Flat Teks
  - Merancangan penyimpanan data dalam flat teks (*plain text*)
- 2. Perancangan Database Relasional
  - Merancangan penyimpanan data dalam konseptual data model
- 3. Perancangan Struktur Menu
  - Merancang menu-menu pada aplikasi yang memiliki fungsi masing-masing sesuai tujuan.
- 3. Perancangan *interface* aplikasi.

Merancang atau mendesain tampilan antarmuka aplikasi dengan pengguna. Tampilan yang dibangun dapat memberikan gambaran umum implementasi dari aplikasi yang telah dibuat.

### 3.6 Implementasi

Pada proses implementasi ini akan dilakukan pembuatan modul yang telah dirancang dan dianalisa selanjutnya diimplementasikan pada bahasa pemrograman dan dilakukan pengujian untuk mengetahui tingkat keberhasilan aplikasi yang telah ada. Berikut adalah spesifikasi lingkungan implementasi perangkat keras dan perangkat lunak :

#### 1. Perangkat keras

Processor	: Intel® Core™ i3 CPU 2.13GHz
Memori (RAM)	: 4,00 GB
Harddisk	: 300 GB

#### 2. Perangkat Lunak

Sistem Operasi	: <i>Windows 7 Home Premium</i>
<i>Tools</i> perancangan	: Notepad++ v6.6.8
<i>Web Browser</i>	: Google Chrome ( <i>Version 37</i> ) / Mozilla 5.0
Bahasa pemrograman	: Hypertext Preprocessor (PHP)
DBMS	: MySQL (mysqlnd 5.0.10)
Perangkat pendukung	: XAMPP 1.8.2, Apache/2.4.9

### 3.7 Pengujian

Pengujian merupakan tahapan dimana sistem akan dijalankan. Tahap pengujian diperlukan sebagai ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan. Pengujian sistem analisis sentimen Twitter ini dilakukan dengan model *confusion matrix* yaitu, sebuah matriks prediksi yang akan dibandingkan dengan kelas asli dari data inputan. Dalam penelitian ini, dibandingkan hasil dari keluaran sistem dengan hasil yang diharapkan dengan permodelan *confusion matrix*.

### **3.8 Kesimpulan dan Saran**

Tahapan ini berisi tentang kesimpulan yang dapat diambil dari penelitian yang telah dilakukan. Pada bagian saran berisi saran-saran yang penulis berikan untuk mengembangkan aplikasi agar kedepannya menjadi lebih baik.